

Récit historique et objet technique : outil de valorisation mutuelle

Matthieu Quantin

IRCCyN, École Centrale de Nantes, CFV, Université de Nantes.

Florent Laroche

IRCCyN, École Centrale de Nantes.

Jean-Louis Kerouanton

CFV, Université de Nantes.

Résumé

Dans le domaine de l'histoire des techniques et du patrimoine matériel, l'informatisation est massive, autant sur le plan de l'écriture, que des sources (archives) ou de la valorisation. Cependant les articulations entre ces plans (archive, recherche, valorisation) restent faibles. Ce travail propose un outil numérique de gestion de contenu historique (récits et sources) pour documenter et valoriser les objets du patrimoine technique. En fonction des situations, différentes interfaces permettent d'accéder aux connaissances, une interface de navigation en 3D est ici proposée en exemple. L'objet technique et le travail de l'historien sont conjointement valorisés.

Le document numérique devient incontournable. Du côté de l'écriture, l'historien ou le conservateur rédige sur ordinateur : récits historiques, renseignements de l'objet, articles, thèses... Du côté des sources, l'objet patrimonial est souvent numérisé (2D ou 3D). De même les références (archives) tendent à être numérisées. Accentuant la tendance, aujourd'hui, notre société produit massivement des documents nativement numériques : sources et objets virtuels de demain. La charte de l'UNESCO entérine cette prise de conscience (UNESCO, 2003). Enfin du côté de l'utilisateur (visiteur), le numérique a fait ses preuves comme outil de médiation, facilitant la compréhension et donc la valeur de l'objet technique.

Aujourd'hui, l'État français encourage les pratiques numériques dans les musées. Les institutions internationales

font de « *l'informatisation [est] un objectif majeur pour la valorisation du patrimoine [industriel]* » (TICCIH, 2003). Les visiteurs de musées plébiscitent ces outils de médiation et les historiens utilisent abondamment les bases documentaires à leur disposition en ligne (Guigueno, 2014). Mais si l'articulation numérique entre sources et recherche est opérationnelle, la valorisation est à part, et semble être une impasse pour la connaissance scientifique, une branche morte. Pourtant Cotte explique en effet que l'usage du numérique « *complète et fait évoluer les méthodes de l'histoire, de l'archéologie, du patrimoine ou de la muséographie, en offrant des possibilités nouvelles de compilation et de mise en scène des connaissances* » (Cotte, 2009).

Face à cela, le projet Nantes1900 (Courtin *et al.*, 2015) mené par notre équipe constituera un repère et servira d'exemple tout au long de cet article pour ses avancées en gestion des connaissances (Hervy *et al.*, 2012). Il est effectivement suggéré de prolonger ce travail par le développement d'outils de Traitement Automatique du Langage pour identifier de nouvelles connaissances candidates à l'intégration dans le système donné (Hervy, 2014, p. 109). Support de réflexion important en termes de gestion de connaissances historiques et de valorisation dans un cadre muséal (appliqué à un objet patrimonial), il est le fruit d'une collaboration entre ingénieurs, historiens et conservateurs. Cinq ambitions en formalisent et prolongent certaines idées directrices.

La première ambition, colonne vertébrale de cet article, est de considérer l'étude historique comme support continu de l'objet patrimonial et non pas seulement comme déclencheur (ou plutôt partie prenante) du processus. Le patrimoine est une construction collective perpétuellement revisitée, et non plus « une fin en soi ». L'idée de support continu est incrémentée d'une dynamique supplémentaire de remise en question permanente, ou plutôt de réévaluation de l'objet. Il s'agit de créer une relation dynamique entre recherches historiques (récits) et objet de patrimoine dans le temps long.

La seconde ambition définit le patrimoine comme bien collectif à se réapproprier. Opposé à l'idée d'un patrimoine transcendantal, l'idée de réappropriation se place du côté de l'usage. Elle établit la valeur d'un patrimoine à partir de la reconnaissance dont celui-ci bénéficie, de sa capacité à être « *continuité entre nous et l'ailleurs d'où il vient* » (Davalon, 2000). Dans une conception topologique du témoignage, « *passé et présent se superposent dans le présent de telle sorte que ce dernier en vient à former en quelque sorte un pli* » (Davalon, 2000) ; il s'agit donc de créer une relation entre visiteur et objet de patrimoine.

La troisième ambition vise l'exploitabilité des collections du musée comme support de connaissances. Les musées possèdent des collections qui sont actuellement difficilement exploitables et dont seule une infime partie (moins de 5 %)

est accessible. Pourtant l'inventaire de ces collections existe.

La quatrième ambition, prolongeant la précédente, consiste à valoriser les collections du musée hors les murs mais aussi dans le musée par la mise en lien d'objets présentés et de collections en réserve (Malraux, 1965). Symétriquement, au vu de la première ambition, ce processus vise aussi à valoriser les connaissances historiques hors des articles.

Enfin la dernière ambition conçoit « *l'article comme API* » (Le Deuff, 2014), afin qu'il devienne possible de requêter¹ un article pour en extraire des connaissances, il faut les structurer et les stocker. Une API désigne une interface de programmation et regroupe un ensemble normalisé de fonctions facilitant la programmation informatique dans le cadre de tâches spécifiques auxquelles elles sont dédiées. Ce dernier point est un prérequis pour l'accomplissement des précédentes ambitions.

D'un côté la production de connaissances est nativement numérique, de l'autre la valorisation *via* le numérique a fait ses preuves. Le développement d'un canal de communication entre ces deux pratiques est une piste à explorer pour tendre vers ces cinq ambitions. L'uti-

lisation du numérique, s'il ne procure pas de solution immédiate et "magique" à ces défis contemporains, peut néanmoins permettre d'envisager un premier ensemble de solutions. Des propositions techniques accompagnées de prototypes fonctionnels prolongent la réflexion entamée avec la récente réalisation Nantes1900.

Notre problématique se déroule comme suit, étant donné la nature massivement numérique de la production des acteurs du patrimoine technique ; comment envisager un enrichissement mutuel entre deux disciplines, la recherche en histoire (des sciences et des techniques) et la valorisation du patrimoine (industriel ou technique) ? Ayant en tête les cinq précédentes ambitions, deux principaux défis seront à relever :

- Du côté de la production, le numérique ne doit pas être un carcan pour l'histoire. Les notions d'incertitudes, temporelles et spatiales ; ainsi que le contexte immatériel de l'objet (« conditions de travail » par exemple) devront être pris en compte et accessibles à la valorisation.
- Du côté de la valorisation, l'accès devra être adapté à l'utilisateur et à la nature des connaissances historiques. Les mêmes contenus ne peuvent pas être présentés au visiteur de passage et à l'expert du domaine.

¹ Abus de langage en informatique, le terme « requêter » fait référence à l'envoi d'une requête vers une base de données pour en extraire du contenu. La requête dépend à la fois du contenu que l'on veut extraire et de la structure de la base de données.

Sources et constats

Tendances actuelles

Deux grands axes permettent de décrire brièvement les tendances actuelles en termes de pratiques numériques dans le monde du patrimoine.

Patrimoine numérique industriel au musée

L'industrie contemporaine produit massivement des données. En restant proche de l'idée d'objet, de maquette, de modèles, de prototypes, tels que conservés dans les musées des techniques, intéressons-nous au cas des modèles 3D produits industriellement. Après les premiers logiciels de « Conception Assistée par Ordinateur » (CAO) volumique, invention attribuée à Pierre Bézier avec UNISURF en France dès 1968, la production numérique s'intensifie rapidement. À partir des années 1980, deux phénomènes participent de sa démocratisation dans le monde industriel : le développement des capacités graphiques pour visualiser de la 3D (à l'écran) et la baisse des coûts, qui deviennent compétitifs face au prix de la main-d'œuvre (dessinateur industriel). Dans cet élan, on observe en France l'industrialisation du logiciel EUCLID en 1980 (issu du CNRS) et d'un logiciel phare du domaine en 1981, CATIA, encore leader du marché aujourd'hui. Aujourd'hui, plus de 30 ans après cette explosion du numérique industriel nous réalisons que ces objets virtuels sont candidats à notre patrimoine industriel. La

décennie 2010 va potentiellement voir l'arrivée programmée de ces objets au titre de patrimoine. En effet, 30 ans cela correspond à la limite basse recommandée par l'Inventaire Général du patrimoine avant la réalisation de l'enquête (Massary & Coste, 2007). Des institutions comme le Cnam devront sans doute prochainement gérer ces cas d'objets techniques nativement numériques.

Accentuant la tendance, le récent essor et la démocratisation des techniques d'acquisition tridimensionnelles et photographiques (couplées à des systèmes de stockage performants) permettent aux musées de numériser leurs collections (scans 3D, photographie HD). Des campagnes de numérisation massives fleurissent dans tous les musées du monde occidental, avec souvent le soutien de l'état, comme au Canada avec le projet Digital Canada 150 ou en France avec différents programmes de numérisation nationaux ou de recherche comme TAPEnAde².

Producteur et utilisateur de données

Parmi les acteurs du patrimoine industriel, les historiens des techniques forment un type particulier de producteurs de données, ils produisent principalement

² TAPEnAde (*Tools and Acquisition Protocols for Enhancing Artifact Documentation*) est un programme de recherche opérationnelle internationale principalement mené et soutenu par des institutions françaises. Lancement du projet en 2011. [URL : <http://www.tapenade.gamsau.archi.fr/>]

des études et des analyses, des récits historiques. Formellement, il s'agit essentiellement de texte. Il est important de noter que ce texte est nativement numérique et son contenu structuré. Des plateformes publiques comme HAL permettent de capitaliser ces articles. L'importance et le dynamisme de la recherche en patrimoine technique et industriel ces dernières années s'est accentué. Formalisé en 2003 par la création du TICCIIH, *The International Committee for the Conservation of the Industrial Heritage*, ce processus s'accompagne d'une considération croissante de la valeur économique du patrimoine industriel (Gasnier, 2007). Dans cette dynamique, le récit en histoire des techniques, l'archéologie industrielle et la monographie tendent à se développer.

Coté utilisateur de données, les enquêtes de satisfaction encouragent les musées à s'équiper de dispositifs numériques. Le rapport Credoc affirme que plus d'un tiers des personnes interrogées utilisent Internet en lien avec la visite d'un musée, d'une exposition ou d'un monument ; dans 16 % des cas il s'agit de visite virtuelle d'une exposition, d'un musée ou d'un monument (Bigot *et al.*, 2012). Par ailleurs les supports de médiation numérique ont su faire leurs preuves auprès du public, notamment dans le cadre de machines industrielles expliquées *via* une maquette CAO comme à Batz sur Mer (Laroche, 2007).

De ce double mouvement, on observe une décorrélation entre les producteurs et les consommateurs : les historiens ne pro-

duisent pas pour les musées et les musées ne « consomment » pas directement les données produites par les historiens. À l'inverse, les historiens exploitent peu les collections de musées dans leurs travaux. Cette étanchéité est décrite à d'autres niveaux, notamment entre recherche et inventaire (Vadelorge, 2014). Ce sont plutôt deux types de données différentes, des systèmes parallèles, deux manières de traiter les mêmes objets sans lien entre elles.

Enfin, dépassant cette opposition entre producteur et consommateur, le grand public produit massivement des données numériques (*crowdsourcing*) et peut participer de la constitution de documentations fiables sur de très nombreux sujets sous certaines conditions (Park, 2014). Ce dépassement de la dualité producteur-consommateur est décrit par le terme « économie de la contribution » et s'inscrit dans un capitalisme cognitif, où la donnée, plus que le produit présente de la valeur. C'est le cas du patrimoine qui n'a de valeur que parce que des études (données) lui en attribuent. Ces concepts sont définis par B. Stiegler (Stiegler, 2009). Le décroisement entre le visiteur et l'objet de collection peut passer par la contribution, nous verrons plus loin quelles formes peut prendre ce mécanisme de contribution.

Cadres existants

Avant de se lancer dans la définition de solutions techniques propres à répondre aux ambitions énoncées, il semble nécessaire de compléter les premiers constats par un aperçu des cadres

existants pour les pratiques numériques dans le monde du patrimoine.

La charte de l'UNESCO (UNESCO, 2003) décrit un ensemble de « bonnes pratiques » relatives à la conservation du patrimoine numérique. Cette charte sera rapidement suivie par des préconisations plus pratiques : la charte de Londres (EPOCH, 2009), qui énonce six principes pour la visualisation assistée par ordinateur dédiée à la recherche et la valorisation du patrimoine culturel. Cette dernière sera étendue par la charte de Séville (*International Forum of Virtual Archaeology*, 2013) dans le cas d'applications à l'archéologie virtuelle. Cependant aucun de ces cadres ne décrit de conseils de mise en œuvre, ni de cahier des charges technique afin de respecter les préconisations d'analyses scientifiques et de capitalisation de connaissances. Sans prétendre à l'écriture d'un tel cahier des charges nous ferons une proposition technique pour le couplage de la médiation de musée avec la rigueur d'analyse historique nécessaire pour le patrimoine.

Concernant les standards de métadonnées, dans le domaine des enregistrements culturels, le choix est large ; mais certaines bases sont incontournables. Deux attentes de bases en termes d'indexation émergent : pouvoir retrouver un item (partie « rangement ») et être capable de créer des liens entre différents items (partie « raisonnement »). Le set de 15 métadonnées DublinCore est inévitable et reprend les bases (date, couverture géographique, auteur...) de manière flexible ;

à l'autre opposé du spectre se trouvent les ontologies (permettant de décrire les relations entre catégories) comme le CIDOC-CRM par exemple. Entre les deux, en fonction des domaines d'application, nous trouvons de nombreux standards : SPECTRUM, CONA, CDWA, LUDO, OAI (sur lequel est basée l'indexation de HAL, la plate-forme française de stockage d'articles scientifiques évoquée précédemment), VRA-Core... (Patel *et al.*, 2005) Enfin, nous estimons que la mise en place d'ontologies n'est pas nécessaire pour créer des liens exploitables entre items. Nous reviendrons sur ce point.

Malgré ces standards et préconisation de bonnes conduites, la tendance à l'entropie est forte. Les cadres sont compliqués à respecter et un clivage entre sciences humaines et sciences pour l'ingénieur limite largement la fluidité de transfert d'informations.

Diversité et constante des pratiques : problèmes soulevés

Des visées à court et long

Au niveau national, en France en 2015, on compte plus de 398 applications mobiles dédiées au patrimoine culturel selon un recensement non exhaustif mené par l'association Club Innovation Culture France (CLIC France)³. Parmi

3 « DOSSIER / 398 applications mobiles muséales et patrimoniales en France (au 28 novembre 2015) » CLIC France.fr [URL : <http://bit.ly/1QgEyb6>]

celles-ci une grande variabilité se dégage, l'usage du numérique en tant que médiation exploite de nombreux moyens techniques : réalité augmentée par des récits et des iconographies, géolocalisation, simulation de fonctionnement mécanique (*via* CAO), audioguide géolocalisé... Cette grande variabilité entraîne un cloisonnement très fort entre les dispositifs développés (ce qui n'est pas vraiment problématique) et une très faible capacité à être améliorés, réutilisés, augmentés... Il s'agit d'une production « jetable ». À l'opposé, du côté des sources, plus centralisées et exprimant des besoins plus stables depuis plus longtemps, l'usage du numérique pour la gestion des bases documentaires est plus rigoureux et s'inscrit dans le long terme.

Cette disparité, entre gestion d'archives sur le long terme et création de dispositifs périssables, complète l'opposition précédemment établie entre producteurs et consommateurs de données historiques. Grâce aux médiums numériques, une transparence et une rigueur accrue dans l'utilisation des données⁴ semblent être envisageables sans surcharger la médiation.

Les projets d'utilisation du numérique au musée à des fins d'analyse, de valorisation et de médiation se multiplient depuis vingt ans (Fleury, 1997 ; Guidi *et al.*, 2008 ; Prévôt, 2013 ; Benoit *et al.*, 2009). Souvent les documents d'archives et la 3D opèrent en synergie pour poser

et répondre aux questions des historiens, mais les livrables finaux sont centrés sur la 3D. La gestion des contenus scientifiques (historiques) de ces dispositifs est peu abordée après la phase d'étude. Seul le projet Nantes1900, pionnier dans la valorisation de contenu historique et d'archives *via* une application muséale s'applique à contextualiser un objet de collection (la maquette). On relève aussi de nombreuses utilisations de la 3D dans le monde du patrimoine et de l'archéologie à des fins d'analyse scientifique et de métrologie. Dans tous les cas, le rendu est toujours découplé des archives et analyses qui sont considérées séparément par le spécialiste (Cassen *et al.* 2014). Ces usages sont bien entendu sources d'inspiration pour ce travail.

Production et transfert de données : l'expérience Nantes1900

La partie valorisation, souvent éloignée de la partie recherche nécessite un important travail de reformatage des données. À ce sujet, le cas du projet Nantes1900 est emblématique. Malgré une forte volonté de documenter et contextualiser un objet de collection à l'aide d'archives et de sources historiques, un travail manuel important a dû être fourni pour faire le lien entre la recherche historique et la valorisation : lecture et reformulation de sources, d'archives, puis (re)formatage pour le processus informatique et validation par le service de conservation. En conséquence de cette série de reformulations, la source historique devient difficile d'accès depuis la valorisation, la traçabi-

⁴ En cohérence avec les Chartes de Londres et Séville.

lité de l'information est compliquée, et donc l'exploitabilité est faible pour les historiens. De plus, le travail manuel de marquage des données est fastidieux, et difficilement exhaustif.

D'un côté, une partie importante du budget et du temps de développement d'une application numérique est dédiée à la recherche de contenus historiques et au processus de validation (pour la médiation) des informations trouvées ; paradoxalement de l'autre, la recherche publique produit d'importantes quantités de connaissances historiques disponibles sous forme de récits. Ceci fut vrai au moins dans le cas du projet Nantes1900. Il n'existe pas encore « d'article API » comme Le Deuff les rêve (Le Deuff, 2014), exploiter des récits historiques est encore un travail peu assisté par ordinateur, presque purement humain.

Le numérique comme carcan de l'histoire

Le numérique souvent sous-exploité est alors perçu comme un carcan pour les sciences humaines.

Une forme de sous-exploitation caractéristique de la rupture entre deux techniques, consiste à plaquer des méthodes de la technique ancienne sur la nouvelle technique. Dans notre cas, l'information sous forme numérique au stade d'article, tout comme au stade de médium de valorisation semble souvent être une transposition de ce qui existait sous forme « non-numérique ». L'écran

prend la place du papier ou du cartel, mais il n'y a pas de personnalisation, l'information est indifféremment redistribuée à tous.

D'un autre côté le numérique, et surtout les représentations 3D rendent souvent l'accès à la connaissance historique dépendant d'un support matériel : modèle physique, reconstitution virtuelle. Il faut combler les trous, l'historien n'a pas (nécessairement) besoin de connaître la forme de chaque pierre ou la pente du toit en tout point pour parler d'une usine. La valorisation par une reconstitution 3D implique d'inventer ces détails. Des informations (géométriques, spatiales, temporelles) parfois très lacunaires ne posent pas de problème réel, souvent même aux yeux de l'historien ces lacunes passent inaperçues. Par exemple, disposer de plans et de descriptions (forcément incomplets) sur les trois quarts des bâtiments d'un quartier industriel à des moments ponctuels, n'empêche pas l'historien d'en dresser le portrait, d'en faire une étude sur la période concernée et de rédiger un récit sur ce quartier. « *Parce qu'il n'existe pas de fait historique élémentaire, d'atome événementiel* » (Veyne, 1971, p. 47) l'exhaustivité est impossible à l'écrit ; en 3D on se contente habituellement de remplir l'espace de représentations souvent inventées et d'y ajouter des informations historiques (on parle souvent « d'annoter ») comme des décorations...

Inversement, il faut aussi « tailler dans la masse », les descriptions issues de récits historiques n'ont pas nécessai-

rement de représentations physiques, il s'agit du contexte immatériel de l'objet. Par exemple « les conditions de travail au début du xx^e siècle »⁵ n'ont pas de lieu précis, elles ne sont pas rattachables à un point d'une modélisation 3D, ni représentables dans l'espace. Il faut bien souvent les éliminer ou les rattacher à des points arbitrairement choisis pour l'occasion. Veiller au découplage entre récit de l'historien et valorisation de l'objet de collection au musée, c'est aussi veiller à leur compatibilité, au « *mapping* » (établir des correspondances) de l'un sur l'autre.

Enfin, les systèmes de gestion de données ne doivent pas établir préalablement des catégories, dans lesquelles les connaissances doivent se ranger, comme les données d'un capteur dans une table de base de données relationnelle. La structure des données doit s'adapter à la connaissance en question, l'inverse serait néfaste à l'histoire et au musée (Srinivasan & Huang, 2005).

En définitive, le découplage à réaliser est multiple, il doit avoir lieu à la fois entre récit de l'historien et l'objet de collection au musée, mais aussi entre le visiteur et l'objet de collection et enfin entre la valorisation et les sources.

Hypothèses de travail

Circonscription de la proposition

À partir des constats précédents et des ambitions évoquées en introduction, nous dressons ici une sorte de cahier des charges de la proposition – ce sont nos hypothèses de travail.

Il est nécessaire d'envisager un pont entre le travail de l'historien et la documentation de l'objet technique patrimonialisé. Ce pont devra permettre la traçabilité de l'information et faciliter l'exploitation historique des collections du musée. Il faudra aussi veiller à la pertinence de l'information transportée du récit historique (monographie) au sein de l'interface tout en minimisant les travaux de reformulation manuels.

La structure des données doit permettre à la connaissance historique de ne pas dépendre d'une représentation physique. Ce dernier point prend en compte le contexte immatériel d'un objet, l'information non situable physiquement sur l'objet, l'information trop imprécise pour établir une forme, un moment ou une position exacte.

Une interface de navigation requête la base de connaissances et permet au public d'accéder aux informations. Cette interface ne doit pas borner l'information historique à l'annotation (par exemple selon le modèle « le saviez-vous ») mais place au contraire l'information historique comme vecteur d'accès.

⁵ Il s'agit du titre d'une fiche de Nantes 1900, compliquée à gérer informatiquement, qui illustre tout à la fois la problématique de spatialisation des connaissances et celle de l'incertitude temporelle en histoire.

Les accès au contenu (c'est-à-dire à l'information transportée depuis les récits historiques jusqu'à l'interface de valorisation) devront être personnalisés en fonction du contexte et en fonction de l'utilisateur. Cette fonctionnalité sera importante pour ne pas submerger d'informations l'utilisateur, ou au contraire le limiter à la lecture de quelques lignes.

La possibilité d'un retour d'informations historiques par le grand public est intéressante (par exemple le *crowdsourcing*) pour l'historien, pour les collections, mais aussi pour la notion de patrimoine en tant que bien à se réapproprier. Ce retour peut prendre une forme volontaire : témoignage, contribution à des collections du musée ; mais aussi involontaire : suivi (« tracking ») de l'utilisateur pour optimiser les visites en fonction des points d'intérêts.

D'un point de vue épistémologique, la modification de la place du récit historique vis-à-vis du patrimoine est fondamentale, le récit n'est plus seulement déclencheur mais accompagne l'objet de musée tout au long de son cycle de vie. Les récits viennent directement enrichir l'objet (ou son contexte) et inversement, l'interaction avec le visiteur permet d'enrichir le corpus du récit historique (cf. point précédent).

Méthodes et outils

À partir des hypothèses de travail précédemment énoncées, nous utilisons des méthodes et outils d'ordre pratiques

et conceptuels. Par la suite nous verrons comment utiliser cette boîte à outils, plutôt hétéroclite.

Le DHRM

L'un des fondements du *Digital Heritage Reference Model* (Laroche, 2007) est le découplage entre interface de valorisation et connaissances historiques en interposant une base de données entre les deux. L'interface se contente de présenter les données contenues dans une base la plus interopérable possible.

Nantes1900 est un grand projet de l'équipe (Hervy *et al.*, 2014), basé sur le DHRM, implémenté dans le musée d'histoire de la ville de Nantes. Ce projet open source est un cas concret de gestion de connaissances associée à un objet de musée. Il s'agit d'un cas éclairant de confrontation des données et méthodes en histoire à la structuration d'un système informatique. Une composante contributive est clairement revendiquée.

Le PLM et la gestion de connaissance

Le *Product Lifecycle Management*, ou « gestion du cycle de vie » en français, est un outil industriel visant à documenter le produit tout au long de son cycle de vie (recherche, conception, développement, production, nouvelle version...). Appliqué au domaine du patrimoine, un tel outil permet de conceptualiser le suivi et la documentation d'un nouveau cycle de vie. L'application de cet outil pourtant bien

implémenté dans le monde industriel n'est pas directe dans notre proposition. La gestion de connaissance est une discipline complémentaire à l'usage du PLM, qui consiste à extraire, mettre en lien et proposer des accès aux connaissances stockées.

Haruspex est un logiciel de gestion de connaissance, issu de l'expérience du traitement de données pour Nantes1900. Spécialement conçu pour traiter les récits historiques et construire une base de données avec indexation supervisée des contenus. Il s'agit de plusieurs modules organisés autour d'un algorithme d'extraction de mots-clefs.

La réalité virtuelle

Les outils de réalité virtuelle et de réalité augmentée, comme les moteurs de jeux, permettent de produire facilement des interfaces et de gérer de la 3D lorsque cela est nécessaire. Nous utilisons un produit classique, unity3D v5.

En termes de méthode, le placement de notre proposition est original (flèche A) : en aval de la production de l'historien. Ce placement contraste avec la plupart des travaux en humanités numériques qui se placent soit en amont et tentent d'intégrer directement avec les pratiques de

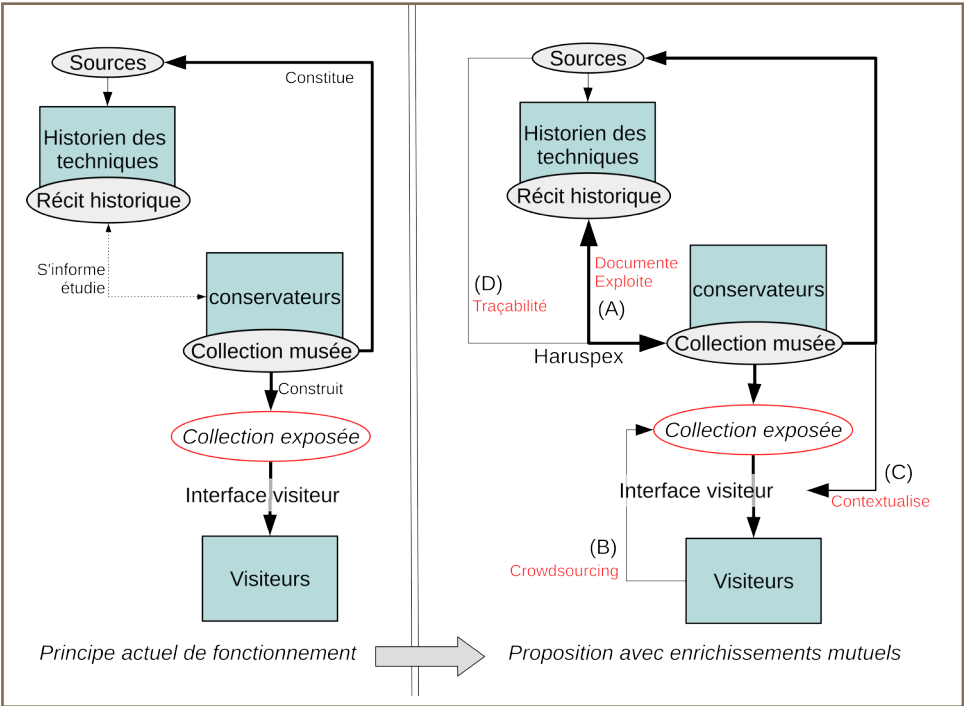


Figure 1 : Schéma des différentes relations aux collections de musée

l'historien, soit très en aval, au niveau des techniques de valorisation du patrimoine mais n'interagissent pas avec l'histoire. L'intérêt de ce placement est de rester au plus près du travail de l'historien d'une part, et d'autre part de ne pas chercher à modifier ses pratiques mais à s'y adapter. La relation de l'historien au patrimoine est plus directe, celle du visiteur moins stéréotypé consommateur (flèche B), mais surtout le visiteur averti peut potentiellement accéder à un véritable « conteur de connaissance » (flèche C). Enfin, une traçabilité des sources documentant l'objet technique est établie (flèche D). D'une certaine manière, comme le montre le schéma de la figure 1, la proposition n'ajoute pas d'acteur mais bouleverse toute la chaîne « classique ».

Développement d'une proposition

Deux parties complémentaires composent la proposition ; d'une part Haruspex permet de capitaliser les connaissances depuis la production de l'historien dans une base de données (orientée graphe), de l'autre une interface compatible avec des objets 3D permet de requêter la base et de visualiser les résultats.

Haruspex

Le principe de fonctionnement de ce logiciel (dans une version antérieure) est décrit (Hervy *et al.*, 2015). En entrée nous avons un texte, et en sortie une base de données graphe composée de « fiches » (morceaux de texte *a priori* cohérents)

liées entre elles par des relations sémantiques pondérées. Le logiciel procède à un découpage du texte selon sa structure (titres, paragraphes...), puis extrait automatiquement les mots-clefs du texte dans son ensemble et indexe les parties entre elles en fonction des occurrences des mots-clefs. Ainsi deux parties où des mots-clefs très similaires apparaissent seront fortement liées dans la base de données. Plusieurs récits linéaires sont transformés en réseau d'informations. Notons que cette nouvelle représentation de l'histoire complète *a posteriori* la forme classique (récit linéaire) mais ne la remplace pas, ni ne s'impose à l'historien. Certaines parties du processus sont supervisées par l'utilisateur : validation et création des mots-clefs notamment.

Ce logiciel tire parti de l'utilisation de contenus historiques *a priori* validés : articles, thèses, mémoires, etc. Une première composante de la dimension *crowdsourcing* fait son apparition, il s'agit de capitaliser de l'information trop souvent cantonnée aux armoires des laboratoires d'histoire. L'article n'est toujours pas une « API » mais il est transformé en contenu exploitable par une machine, « requêteable ». Ce module cherche à renseigner des documents non codifiés à actualisation variable dans un système codifié à actualisation fréquente selon le système établi par Zacklad (Zacklad, 2007).

Comme l'opérationnalité du projet Nantes1900 nous l'a démontré, l'application ne doit pas devenir un prétexte à mettre en œuvre un système documen-

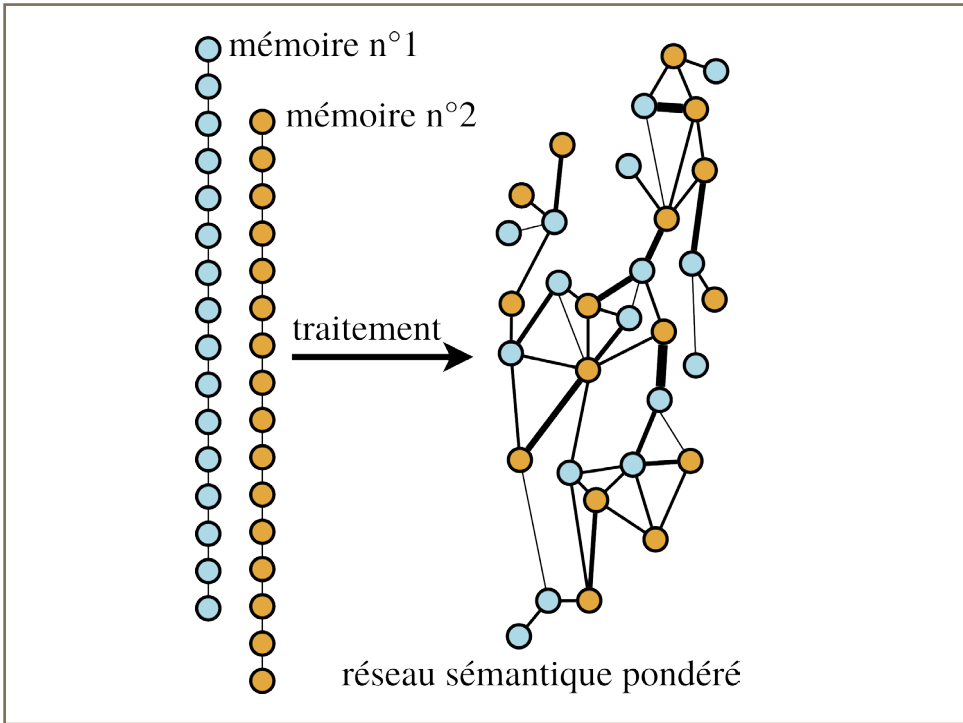


Figure 2 : Construction du réseau à partir de fiches issues de récits linéaires

taire d'une complexité trop importante par rapport au besoin. Ainsi nous éviterons l'usage d'ontologies ou de systèmes de descripteurs externes (thésaurus).

À partir de ces connaissances historiques découpées et rangées, les données acquises sont structurées à l'aide de descripteurs simples et couvrant l'ensemble de la notion « d'information historique ». Ces descripteurs sont temporels, sémantiques et géographiques. Ils recoupent les types de base du Dublin Core. Toutes les informations historiques sont « situables » quelque part dans l'espace décrit par les trois axes du schéma ci-

dessous. Le corollaire de ce postulat est le suivant, une information historique qui n'a ni composante thématique, ni spatiale, ni temporelle n'a pas d'existence à nos yeux. Cette hypothèse est jugée vraisemblable.

Le type des métadonnées Dublin Core n'étant pas fixé, nous devons utiliser un formalisme simple pour permettre la comparaison et l'interopérabilité de ces données. Ainsi, pour les dates nous utilisons le standard ISO 8601 et des schèmes de vocabulaire contrôlé. Pour les lieux nous utilisons un vocabulaire contrôlé ; des zones sont définies et nommées,

elles correspondent aux entités géographiques utilisables. Ces deux derniers éléments génèrent des problèmes de prise en compte des incertitudes temporelles (Leduc, 1999). Dans un souci de simplicité et pour une première approximation, nous sommes contraints de réduire « début de xx^e siècle » à l'intervalle 1900-1950, ou « Salons Mauduit » à une zone précisément bornée. Ceci implique un travail manuel et un temps de définition préalable.

La partie sémantique est entièrement gérée par Haruspex, les mots-clefs les plus spécifiques constituent le vocabulaire contrôlé caractéristique d'un récit et les mots-clefs les plus génériques servent à lier les éléments entre eux : différents récits, 3D, iconographie... La conformité du réseau sémantique obtenu est au plus proche du récit linéaire d'origine. Cette approche s'oppose à l'indexation de texte dans des boîtes préétablies, dans ce dernier cas on parle plutôt de « reconnaissance »⁶ que d'extraction.

La contrepartie de ce mécanisme est l'absence de vocabulaire contrôlé pour l'indexation sémantique. Pour pallier ce manque, deux stratégies sont à l'œuvre. La première consiste à construire le vocabulaire de manière supervisée (l'utilisateur peut modifier les mots-clefs extraits). La seconde consiste à associer plusieurs formes syntaxiques à chaque mot-clef ; l'une de ces formes est une « bonne » réponse à une requête

vers des ensembles de données externes, vers DBpedia (Mendes *et al.*, 2012) ou dataBNF par exemple. Ainsi nous nous rattachons potentiellement à une forme normalisée. Ces références externes sont communes et conformes aux standards en web sémantique notamment. Mais cette composante « tournée vers l'extérieur » n'est pas l'objet de cet article. Dans tous les cas la forme spécifique au récit historique est conservée.

Face à l'enjeu de la traçabilité de l'information et des sources historiques, Haruspex se charge de conserver la trace des références telles que cités par l'historien, auteur du document d'origine. La gestion des citations au fil de la narration (*op. cit.* ; *ibidem* ; *ibid.*...) et la normalisation des résultats pour éviter les doublons sont prises en charge. Le pointage de références, depuis une interface de valorisation vers des sources stables, disponibles sous forme numérisée si possible, gérées par des services d'archives se présente comme une solution pour permettre la réutilisation des données et la traçabilité de celles-ci.

Les métadonnées utilisées sont rudimentaires et se combinent facilement, comme nous venons de le voir, toute information historique est descriptible par au moins un des trois axes (espace, temps, thématique). Cette simplicité vise à maximiser l'interopérabilité d'un récit à l'autre (créer des connexions entre les récits) tout en évitant les systèmes complexes. En effet l'ajout d'informations issues de nouveaux récits, l'enrichissement de la base au cours du temps, né-

⁶ L'acronyme consacré est NER(C) : *Named Entity Recognition (and Classification)*.

cessite une compatibilité maximale. De plus cela prévoit la réutilisation possible des données pour de futures utilisations, ou des utilisations parallèles (qui peuvent qualifier davantage les données si nécessaire). Néanmoins ces descripteurs semblent suffisants pour créer des liens, comme nous allons le voir de suite.

Pour chaque lien entre deux items, trois valeurs sont associées : une pour la proximité spatiale, une pour la concomitance temporelle et une pour la similarité thématique. Une première pondération

globale du lien est obtenue en combinant ces trois valeurs. À cette pondération viennent s'ajouter les fréquentations du lien (si tout le monde clique au même endroit alors cela doit être intéressant). Enfin, l'administrateur peut forcer manuellement la pondération du lien, associant ou dissociant fortement un item à un autre, quel que soit le résultat des calculs de proximité.

L'avantage de cette modélisation réside dans sa simplicité (pour le conservateur), tout en permettant une analyse (par l'historien) et une valorisation (pour

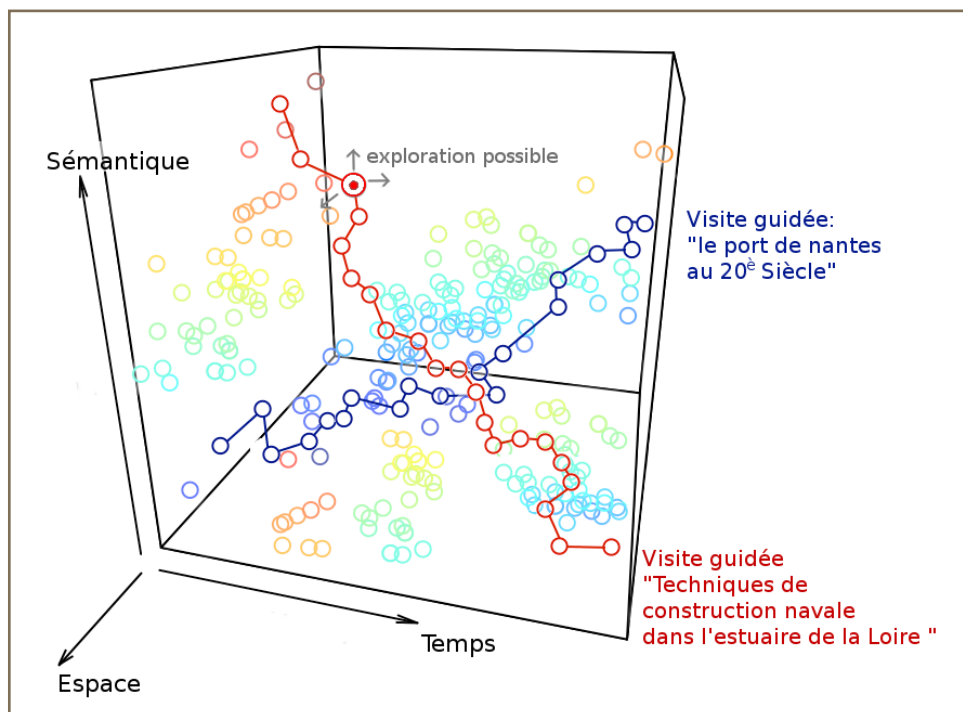


Figure 3 : Schéma présentant 2 visites guidées : chemins frayés parmi les données en 3 dimensions (espace, temps, thématique).
À chaque point de la visite, il est possible de s'éloigner du "fil conducteur" pour explorer les données proches

le public) pertinente. En effet il est possible de créer des visites guidées en « balisant un chemin » (liens forcés) dans ces trois dimensions (temps, espace, thématique), et de suggérer une exploration basée sur la proximité spatiale, temporelle, thématique (voir schéma figure 3).

Une telle structure de données ne permettra pas à la machine de raisonner : pas de déduction du type « deux personnes appartiennent à la même entreprise donc ils sont collègues », pas de création automatique de visite guidée, ni de rétrocréation de récit historique (association automatique d'items en un récit). Nous considérons que le travail du médiateur et de l'historien est proprement humain et irremplaçable. Il s'agit donc d'un outil d'aide à la décision pour les professionnels. Voyons maintenant en quoi cet outil présente un potentiel fort pour la relation au public, comment sont gérés les problèmes de représentation des données non spatialisées évoqués précédemment.

Visite guidée mais libre

L'aspect valorisation de la proposition se présente sous la forme d'une interface d'accès aux connaissances précédemment capitalisées. Comme décrit dans les hypothèses de travail, cette interface ne doit pas « brider » les connaissances stockées, pas de censure, pas de surinterprétations pour des besoins de représentation géométrique ou esthétique. Nantes1900 présente quelques cas d'informations non situées mais accessibles *via* l'interface. Certaines fiches n'ont pas de représentation sur la

maquette du port industriel de Nantes en 1900 et sont uniquement accessibles *via* d'autres fiches. Il s'agit là des prémisses de l'idée que nous essayons de poursuivre en décentrant l'interface jusqu'alors basée sur des représentations physiques (la maquette est au centre).

Toujours dans la démarche de recherche d'information de Nantes1900, sur le principe « *overview first, zoom and filter, then details on demand* » (Shneiderman, 1996), nous recentrons l'accès sur l'information. Pour éviter que les connaissances historiques ne dépendent d'une représentation, nous avons fait le parti de construire cette interface comme un moyen d'accéder à des représentations *via* des descriptifs abstraits, propres à l'information. Cette démarche « à contre-pied » de ce qui se fait habituellement, utilise les descripteurs espace – temps – thématique pour naviguer dans la masse documentaire. Nous accédons aux connaissances (et éventuellement à leurs représentations) *via* des descripteurs. La plupart des informations sont associées à des représentations : iconographie, modélisation 3D, situation géographique, mais cela n'est pas nécessaire. Concrètement, l'interface propose à l'utilisateur des informations connexes à ce qu'il consulte. Il est possible de moduler ce niveau d'exploration (affichage des données connexes) par des curseurs de degré de proximité pour chacune des trois variables (espace, temps, proximité sémantique). L'interface affichera uniquement les informations très liées par cet ensemble de critères pondérés, ou au contraire l'ensemble des informations (même faible-

ment) associées à l'information consultée. De même, l'accès aux sources telles que référencées par le récit historique initial assure la traçabilité de l'information et permet l'exploitation de celle-ci par les historiens. L'interface permet différents accès à l'information, ainsi les sources ne sont pas imposées mais disponibles. Par la génération de liens forcés de médiation (visite guidée), un chaînage des connaissances associées à une scénarisation permet de générer des parcours statiques pour les utilisateurs qui ne désirent pas s'aventurer dans l'ensemble de la masse documentaire.

Ce mécanisme est proche de la conception d'une visite guidée classique, où un guide vous emmène pas à pas dans une péripiétie à travers divers thèmes, époques, espaces. Nous prenons ici un parti différent de celui de Nantes1900 qui a définitivement opté pour un dispositif de

médiation libre, sans guide. Dans notre cas, la visite peut être interrompue à tout moment pour laisser l'utilisateur s'aventurer dans un sujet connexe.

Cette interface mêle donc représentations physiques et abstractions, visite libre et visite guidée. Capable de proposer un parcours préétabli par le service de médiation, tout en ouvrant l'accès au réseau de connaissances dans toute sa complexité, elle s'adresse différemment au visiteur, au curieux ou à l'historien spécialiste du domaine en question.

Cette interface n'est qu'un prototype visant à démontrer les capacités d'une telle conception de la médiation et de la gestion des accès à la connaissance historique. Une implémentation plus ergonomique et appliquée à un cas concret est actuellement en développement.

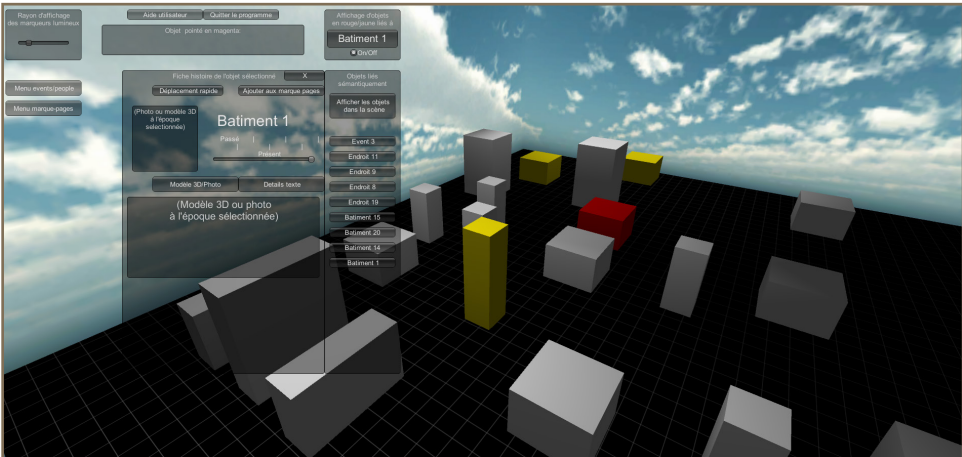


Figure 4 : Prototype d'interface permettant d'accéder *via* des descripteurs simples (proximité spatiale, temporelle et sémantique) à des représentations de connaissances historiques (texte, iconographie, modèle 3D) liées entre elles

Le terme « accès aux données » inclut à la fois la consommation de données mais aussi la production, la contribution, ce que nous avons nommé le crowdsourcing. Dans ce cas précis des Salons Mauduit de Nantes, la composante populaire de ce petit patrimoine a été jugée essentielle. Pour cela, une interface de dépose de documents d'archives privées (à l'instar du projet Nantes1900) et de témoignages a été mise en place. La simplicité des descripteurs précédemment évoquée (espace-temps-thématique) facilite la dépose de documents et de témoignages, tout en maintenant une cohérence avec l'existant. Cette étape est nécessairement modérée par des spécialistes du domaine.

Résultats et perspectives

Plusieurs résultats sont à évaluer. En effet, ce projet est composé de deux principaux modules et ayant une portée épistémologique importante. Nous choisissons d'évaluer ces différentes composantes séparément. De plus, le cadre applicatif à un prototype opérationnel les « Salons Mauduit » en tant qu'expérience prolongeant le travail effectué avec le Musée d'Histoire de la Ville de Nantes (Nantes1900) mérite une évaluation spécifique.

Résultats pour le module Haruspex

Ce module, qui construit des lectures multi-points-de-vues répondant à un besoin méthodologique exprimé par les historiens

(Cotte, 2007), est une première brique pour la valorisation des connaissances. Les capacités d'agrégation de contenu du système permettent d'étendre le réseau d'informations, et d'en suivre toutes ses évolutions, sans limite, à condition que l'information soit représentable par les descripteurs temps-espace-sémantique.

Ce système est flexible et évolutif. Fondé sur des descripteurs simples, il permet l'agrégation de données très hétérogènes. Considérant que l'objet technique entre dans un nouveau cycle de vie lorsqu'il est patrimonialisé, et que l'agrégation de nouvelles informations est un facteur de variabilité dans le temps, on saisit le lien à la notion de « gestion du cycle de vie ». En effet, l'analyse évolue au cours de la vie de l'objet et nous croyons que ces différentes versions ne sont pas destinées à se substituer les unes aux autres mais à se compléter dans une agrégation de contenus issus de divers points de vue. Les différentes versions qui se complètent décrivent une réalité, un fait historique au sens où Paul Veyne l'entend (Veyne, 1971). Aussi, nous insistons sur la notion de cycle de vie opposée à un état figé. Haruspex permet d'un côté la mise à jour des informations relatives à l'objet qui évolue au fil des recherches historiques et des contributions (sources, témoignages) et de l'autre la confrontation des différents points de vue. Cette notion de cycle de vie patrimonial de l'objet technique est directement issue du monde de l'industrie où elle a su faire ses preuves⁷. Inversement, nous

⁷ On parle alors de PLM pour *Product Lifecycle Management*.

estimons que la gestion de la complexité de l'objet patrimonial peut apporter des solutions au monde industriel.

La possibilité de manipuler des connaissances abstraites, immatérielles et d'agréger plusieurs sources connexes construit naturellement un contexte à l'objet technique. Situé par ses composantes « temps, espace, thématique », l'objet est en lien avec son voisinage. Ainsi les autres items proches constituent le contexte de l'objet. On peut parler de contextualisation mutuelle. Une situation faiblement contextualisée correspond à

un manque d'information, à un maillage d'informations ténu, ce qui se traduit sur le graphique (Figure 3) par une zone à faible densité de points.

Concernant la gestion de l'imprécision, nécessaire en histoire, le problème n'est réglé qu'en apparence, tronquant les résultats et laissant à l'utilisateur une part d'interprétation. Les données spatiales sont simplifiées à des zones associées à un mot (« grande salle » par exemple). Les données temporelles peuvent être non seulement représentées par des dates mais aussi par des mots

The screenshot shows the Haruspex application window with the 'Pré-traitement' tab selected. The interface includes several input fields and checkboxes for configuring document processing parameters.

Field/Option	Value/State
Dossier du projet	[Empty text field] [Parcourir]
Découpe des paragraphes	<input type="checkbox"/>
Fermeture des crochets	<input type="checkbox"/>
Taille minimum des fiches (en mots)	150
Écrire la dernière section? (toujours vrai si plusieurs fichier)	<input type="checkbox"/>
Auteur	Bertrand Dumas
Date de publication	février 2016
[Enregistrer les paramètres]	
[Créer le fichier]	

Figure 5 : Capture d'écran de l'interface de commande d'Haruspex

(nom d'une époque par exemple). Dans les deux cas ce vocabulaire restreint dédié à l'espace et au temps est contrôlé et chaque mot (époque ou lieu) correspond à un ensemble fixe et stricte. En ce sens la gestion de l'écueil de l'incertitude temporelle et spatiale est évitée, mais le problème persiste dans le fond. Derrière le flou apparent du mot (époque, lieu) pour l'utilisateur, se cache une inflexible exactitude (dates, coordonnées de position) pour la machine.

Une attention particulière a été portée sur le respect du texte historique dans sa spécificité. La création de boîtes issues du texte (les mots-clefs extraits par Haruspex) pour son indexation minimise les approximations et assure l'adéquation des descripteurs sémantiques au contenu indexé. La possibilité de superviser les résultats assure des résultats de qualité comme nous allons le voir ci-après. La normalisation des descripteurs se fait *via* des requêtes à DBpedia. Le concept de standard est ici utilisé comme une surface d'échange et non pas comme un ensemble de descripteurs hégémonique. Ainsi le système proposé ne cherche pas à bouleverser et remplacer les standards existants (CIDOC-CRM, SPEC-TRUM, VRA-CORE), mais plutôt à les compléter localement pour l'analyse et la valorisation. À des fins d'échange de données il faut renoncer à la spécificité de notre objet, de nos collections, de notre documentation. Haruspex conserve cette spécificité localement, isolément du « *fantasme récurrent de la bibliothèque universelle* » (Welger-Barboza, 2001).

Évaluation sur corpus

Plusieurs tests de fiabilité ont été effectués pour Haruspex. La mesure appelée « rappel » concerne les mots manquants pour l'indexation d'un texte, ceux qui n'ont pas été trouvés par la machine, mais évalués selon l'auteur du texte. La mesure de « précision » concerne les mots trouvés qui sont erronés, ceux qui ne représentent pas le texte, évalué selon l'auteur encore. En moyenne sur des corpus de 1 à 39 documents, comportant entre 20 et 300 mille mots, nous obtenons 72 % de précision, et 98 % de rappel. La note composée de ces deux mesures, appelée score F1, est alors de 0.83 ce qui est dans la moyenne. L'intérêt principal d'un processus supervisé est d'améliorer la précision en facilitant la suppression manuelle des extractions erronées⁸. Cet algorithme permet l'extraction de mots-clefs extrêmement spécifiques comme « maçonnerie de moellons hourdés de chaux », « chambre syndicale des constructeurs de navires » ou encore « turbine à réducteur à engrenages ».

Quelques chiffres, enfin : environ 300 mots-clefs sont extraits pour 22 mille mots de texte, 650 pour 67 mille mots et 1 000 pour 300 mille mots. Environ 230 mille liens sont créés pour 1 000 mots-clefs extraits. Les textes utilisés pour ces évaluations sont des mémoires de Master 2 et Thèses en histoire ou histoire des

⁸ Ceci permet d'améliorer la précision et d'orienter les résultats *a posteriori*.

techniques, des recueils d'entretiens en laboratoire, des corpus d'archives⁹.

Perspectives pour Haruspex

Haruspex est en développement et les résultats intéressants mériteraient d'être améliorés sur certains points. La prise en compte des incertitudes est un point étudié par ailleurs et peu intégré dans notre gestion des données temporelles et spatiales. L'utilisation de la logique floue semble notamment prometteuse (Desjardin *et al.* 2012). Un second point d'amélioration concerne la normalisation des citations et références extérieures extraites du texte pour éviter les doublons. Néanmoins la traçabilité de l'information, objectif le plus préoccupant, est atteinte.

Résultats pour l'interface utilisateur

Deux notions importantes issues de Haruspex concernent l'interface utilisateur : la création de liens typés et la pondération de ces liens. Il s'agit de lier des documents entre eux selon différents critères (typage) : temps, espace, thématiques et d'attribuer un coefficient (pondération) à ce lien indiquant sa « force » sur tel ou tel critère. Ces notions permettent de concevoir des niveaux de lecture, une personnalisation de l'accès à l'information.

Ce module de la proposition se place moins sur le plan de l'apport scientifique pour l'historien que pour éviter l'effet « impasse » de la médiation et de la valorisation dans les musées. Il s'inscrit dans une réflexion sur la place des représentations au sein du Consortium 3D de la Très Grande Infrastructure de Recherche (TGIR) Huma-Num et constitue en quelque sorte la partie émergée de notre travail. Il n'y a malheureusement pas eu d'évaluation utilisateur de ce système. Nous sommes conscients de son insuffisance ergonomique pour une réelle implantation (en l'état) dans un musée. Cependant ce démonstrateur (Quantin, 2015) prouve qu'il est possible de représenter des connaissances abstraites et de créer ainsi un canal de communication dynamique entre le récit historique et l'objet de patrimoine technique.

La personnalisation des accès tire profit du numérique au sens où l'information présentée est modulable : de la visite guidée à l'accès, à tous les niveaux les ressources sont liées pour une exploration en profondeur. La représentation figée, directement issue de la culture du papier, est dépassée mais n'est pas abolie. En effet la visite guidée et le récit historique classiques peuvent continuer à exister conjointement.

Pour les musées, ce système permet la conservation et la valorisation du patrimoine immatériel en tant que contexte d'un objet de collection, sans pour autant surcharger l'utilisateur. L'objet de collection devient alors un conteneur de

⁹ Parfois *via* un process de scan puis OCR (*optical character recognition*), c'est-à-dire obtention d'un texte brut depuis l'image d'une feuille de texte.

connaissance. Le projet Nantes1900 a été bâti avec cette même intention. Il a permis d'ouvrir la voie et d'approfondir l'idée d'un objet de patrimoine révélateur et représentatif d'un ensemble grâce au numérique.

Héritant des concepts du DHRM et de l'expérience réussie de Nantes1900, l'information est indexée dans une base et pointe vers des ressources stables. Le dispositif s'inscrit dans le long terme et sa péremption se limite à l'obsolescence de l'interface (les données ne périssent pas). La même base peut servir à des usages très variés, l'interface peut être remise « au goût du jour », les usages peuvent se diversifier en parallèle les uns des autres, le mode ludique par exemple n'a pas été abordé.

Ce type d'interface présente les collections et les liens qu'elles tissent entre elles et avec des références extérieures. Le musée s'émancipe du coût de reconstitution historique multimédia par des prestataires, avec une approche scientifique parfois douteuse et une pérennité très faible. L'institution valorise ses collections en réserve et affirme son rôle d'étudier, conserver, exposer et transmettre le patrimoine matériel et immatériel (ICOM, 2007).

Une seconde interface permet de déposer des documents, de témoigner et consulter les documents en ligne. Il s'agit d'une interface web simple et mettant à disposition le contenu. Basé sur le Système de Gestion de Contenu web (CMS) Omeka

(Roy Rosenzweig Center for History and New Media, 2016), dédié à l'affichage de certaines données, à la collecte et à l'administration des contributions et témoignages des utilisateur-acteurs : témoin, visiteur averti ou historien. Ces derniers cèdent leurs droits sur ces contenus.

Perspective et limites pour cette interface

Basé sur la contribution, principalement d'origine scientifique, le système dépend des droits à utiliser et diffuser cette connaissance, de la liberté des historiens à partager leur savoir. Cette problématique nous ramène à la question de bien commun informationnel, à la possibilité d'utiliser librement la production scientifique (Hess & Ostrom, 2003, chap. 5) en histoire à des fins muséales. Il s'agit là d'une question fondamentale pour la viabilité du système proposé. Dans ce même registre, le projet Nantes1900 a connu quelques difficultés et a renoncé pour l'instant à diffuser les contenus extraits de services externes comme les Archives Municipales de Nantes et Départementales de Loire Atlantique.

L'amélioration de la personnalisation des accès est un axe de développement intéressant. L'objectif serait non plus de proposer des items en lien avec ce qui est consulté mais de construire des trajectoires et de réduire l'éventail des suggestions à chaque étape, en fonction du parcours qui a mené l'utilisateur à consulter tel ou tel item. De

plus, le développement d'une interface plus soignée est prioritaire, non pas à des fins « cosmétiques » mais plutôt de support pour davantage de collaboration interdisciplinaire en lien avec le musée (Mason, 2015).

La possibilité de géolocaliser un utilisateur *via* son terminal (*smartphone*) et de lui fournir les informations relatives à la zone dans laquelle il se trouve est une piste que nous explorons actuellement pour sortir la connaissance des musées, et la rendre accessible hors les murs. Ceci est particulièrement pertinent pour le patrimoine technique et industriel, où les objets sont volumineux. Dans le cas des Salons Mauduit nous explorons sérieusement cette possibilité avec l'objectif de mettre en lien des objets de collections

extérieures, des sources historiques et des objets physiquement présents devant l'utilisateur. On retrouve là l'idée du musée hors les murs et de la connaissance historique hors des articles évoqués dans les ambitions, en introduction. Des outils développés autour du projet Européen TagCloud sont intéressants à ce titre (Floch *et al.*, 2014), mais orientent fortement la médiation autour de l'expérience utilisateur, parfois au détriment du contenu scientifique.

Couplage

Face à la difficulté pour un musée de gérer les contributions des utilisateurs, comme c'est le cas pour le projet Nantes1900 qui réfléchit encore à la



Figure 6 : Maquette de l'interface de navigation entre les documents selon différents critères, temps, espace et sémantique. Possibilité de suivre une visite guidée.
Repère géographique local *via* le scan 3D du lieu

meilleure façon de procéder et valider ces informations¹⁰, la volonté d'agir s'est portée sur un ensemble de plus petite ampleur : le projet de Salons Mauduit. Ce projet qui concerne un patrimoine local nantais, en partenariat avec l'Université de Nantes et la ville de Nantes, s'intéresse notamment à la dimension populaire du lieu. Une amorce de 180 documents d'archives, constitue un terrain d'expérimentation pour le couplage du module Haruspex et de l'interface de valorisation présenté dans cet article.

Ce lieu est en travaux de reconstruction, et un dispositif de valorisation *in situ* est prévu pour la fin du chantier. Cette salle de réception art déco a reçu trois générations de Nantais, de tous horizons, et cristallise beaucoup de souvenirs locaux. La visite d'un bâtiment se prête bien à entremêler patrimoine matériel et immatériel. Cela constitue donc un cas d'étude adapté.

L'interface précédemment proposée doit être adaptée pour ce cas d'histoire d'un bâtiment. Le projet étant en attente de validation d'un jalon, l'interface est pour l'instant plus présentable sous forme de maquette.

Le module de calcul de distance sémantique, temporelle, et spatial fonctionne, ainsi chaque document est lié à un ensemble d'autres, plus ou moins

fortement. Des documents qui n'ont pas de localisation se trouvent fortement liés à d'autres documents situés, ce qui en permet l'accès. Comme le montre la figure 6, nous avons opté pour la mise en situation avant reconstruction des lieux. Pour cela un scan3D intégral de l'intérieur du bâtiment a été effectué. Néanmoins il est possible d'accéder à des informations concernant des zones non scannées (l'extérieur par exemple), ou non reconstruites mais scannées (les cuisines par exemple).

Épistémologie et perspectives

La mise en relation de textes à l'origine conçus comme des récits linéaires parallèles, rejoint l'idée de typicité du patrimoine. Cette idée est décrite par (Heinich, 2009) comme étant constitutive d'objet de patrimoine, à l'opposé de l'unicité. L'objet est alors conçu comme représentant d'une série, comme élément typique d'un ensemble d'objets. Illustrons comment la proposition s'inscrit dans cette idée, en restant à Nantes et choisissant un exemple caractéristique du monde industriel, le pont roulant. Plusieurs sources historiques concernent la construction navale. Un de ces ouvrages s'intéresse particulièrement à la place du pont roulant dans cette industrie, d'autres l'évoquent. Aujourd'hui des travaux de réhabilitation modifient un grand ensemble industriel de l'île de Nantes, les « Halles Alstom ». La question des ponts roulants a été abordée, et ils furent considérés comme représentant typique de l'activité industrielle. Certains

¹⁰ Xavier Gerard, « 7 questions à Christophe Courtin ». Club-innovation-culture.fr. 2015 [URL : <http://bit.ly/1JOmMhw>] consulté le 1^{er} février 2016.

d'entre eux seront conservés. Demain, dans les bâtiments réhabilités, un dispositif de médiation pourra mettre en lien les ponts roulants rescapés avec toutes les fiches contenant des références au pont roulant ; ce dispositif sera transversal à plusieurs études historiques et documents d'archives. Ce dispositif proposera de documenter le pont roulant ou d'en témoigner. Cet objet deviendra le représentant d'une série de ponts roulants et d'un ensemble de connaissances associées (logistique, dimensions des halls, évolution des techniques de production, etc.). Cette possibilité découle du fait qu'aucun *a priori* ne définit l'indexation des récits. Le mot « pont-roulant » peut émerger au croisement d'un grand nombre de récits distincts et ne traitant pas exclusivement ce sujet à l'origine.

L'impact direct de ce système sur les techniques classiques de l'historien (rédaction de récit linéaire) est nul, c'est-

à-dire qu'il n'influence pas les pratiques classiques de l'historien des techniques. En effet il se situe en aval de cette production et propose une représentation des connaissances complémentaire à l'article. De même pour le conservateur de musée. En termes de métier, la seule modification notable à prendre en compte est la nécessaire modération du *crowdsourcing* et le suivi de l'indexation supervisée, cette dernière tâche tend à être optimisée pour être la plus légère possible.

Conclusion

Le développement de ce système à deux têtes, d'un côté tourné vers la production des historiens des techniques, de l'autre vers le musée et le visiteur, démontre le potentiel qu'il existe à relier les pratiques des historiens et les besoins du musée. Ce système met en évidence les

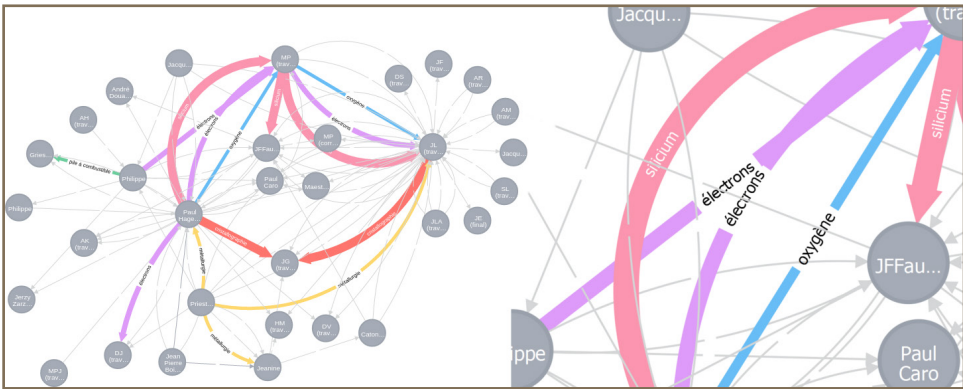


Figure 7 : Visualisation graphique (et gros plan) du retour d'une requête concernant les liens sémantiques les plus forts entre quelques nœuds d'un graphe de fiches sur les pratiques de la chimie en laboratoire. Obtenu via Haruspex...

défauts existants et tente d'y remédier : difficulté à exploiter les récits historiques directement, coût de recherche pour un dispositif de médiation en musée, *a priori* incompatibilité entre données historiques incertaines ou lacunaires et accès interactif pour le visiteur de musée. De futurs développements semblent nécessaires pour répondre aux perspectives d'amélioration évoquées précédemment. L'application à un grand nombre de cas pratiques permettra sans doute de mettre en évidence d'autres manques, cependant ces outils existent concrètement aujourd'hui, ce système fonctionne et s'inscrit dans une dynamique d'amélioration de projets d'envergure déjà réalisés par notre équipe comme Nantes1900.

Bibliographie

- Benoit S. *et al.* (2009). « Usines 3D. La simulation pour questionner les sources et les vestiges de l'histoire industrielle ». In Vergniew R. & Delevoie C. (éd.), *Virtual Retrospect 2009*. Pessac : Ausonius, pp. 31-40.
- Bigot R. *et al.* (2012). *La visite des musées, des expositions et des monuments*. Étude pour la Direction Générale des Patrimoines.
- Guigueno B. (2015). *Qui sont les publics des archives ?* Étude pour la Direction générale des patrimoines.
- Cassen S. *et al.* (2014). « Complementarity of acquisition techniques for the documentation of Neolithic engravings : Lasergrammetric and photographic recording in Gavrinis passage tomb (Brittany, France) ». *Journal of Archaeological Science*, 45(1), pp. 126-140.
- Cotte M. (2007). « La génétique technique a-t-elle un avenir comme méthode de l'histoire des techniques ? » In Rey A.-L. (dir.), *Méthode et Histoire, journées d'études de la SFHST*. Lille : publications de la SFHST, pp. 187-201.
- Cotte M. (2009). « Les techniques numériques et l'histoire des techniques. Le cas des maquettes virtuelles animées ». *Documents pour l'histoire des techniques*, 18, pp. 7-21.
- Courtin C. *et al.* (2015). *Nantes 1900 – La maquette du port. Une valorisation scientifique et muséographique innovante*. Nantes : Musée d'histoire de Nantes.
- Davallon J. (2000). « Le patrimoine : "une filiation inversée" ? » *Espaces Temps*, 74(1), pp. 6-16.
- Desjardin É., Nocent O. & de Runz C. (2012). « Prise en compte de l'imperfection des connaissances depuis la saisie des données jusqu'à la restitution 3D ». In Giligny F. & Costa L. (dir.), *Archeologia e Calcolatori*, p. 389.
- Le Deuff O. (2014). « Le temps des changements ». In Le Deuff O. (dir.), *Le temps des humanités digitales*. Limoges : FYP, pp. 115-172.
- EPOCH (2009). *London Charter*. Site Web [En ligne].
- Fleury P. (1997). « La Rome Antique sur l'Internet ». *Revue Informatique et Statistique dans les Sciences Humaines*, 33, pp. 146-162.
- Floch J., Jiang S. & Beltrán M. E. (2014). « Tailoring Lifelong Cultural Experiences ». In Stephanidis C. & Antona M. (dir.), *Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge, (part 2)*, pp. 681-692.
- Gasnier M. (2007). *Le patrimoine industriel comme vecteur de reconquête économique*. Belfort : Lavauzelle.
- Guidi G. *et al.* (2008). « Virtualising

ancient imperial Rome : from Gismondi's physical model to a new virtual reality application ». *International Journal of Digital Culture and Electronic Tourism*, 1(2), pp. 240-252.

Heinich N. (2009). *La fabrique du patrimoine, de la cathédrale à la petite cuillère*. Paris : Maison des Sciences de l'Homme.

Hervy B. *et al.* (2014). « Augmented historical scale model for museums : from curation to multi-modal promotion ». *Laval Virtual VRIC 14*. Laval (France), pp. 10-13.

Hervy B. (2014). *Modélisation et intégration de données hétérogènes sur un cycle de vie produit complexe*. Thèse de doctorat soutenue à l'École Centrale de Nantes.

Hervy B., Laroche F. & Bernard A. (2012). « An Information System for Driving the Future PLM for Museum : The DHRM, Digital Heritage Reference Model ». *Advanced Composite Materials and Processing ; Robotics ; Information Management and PLM ; Design Engineering*, pp. 465-471.

Hervy B., Quantin M. & Laroche F. (2015). « Gestion de connaissances pour l'acquisition, le traitement et la valorisation des connaissances du patrimoine technique ». *Ingénierie des connaissances – IC2015*. Rennes.

Hess C. & Ostrom E. (2003). « Ideas, Artifacts, and Facilities : Information as a Common-Pool Resource ». *Law and Contemporary Problems*, 111(1998), pp. 111-146.

International Forum of Virtual Archaeology (2013). *Principles of Seville. International Principles of Virtual Archeology*. Seville.

ICOM (2007). *Statuts du Conseil international des musées*. Autriche.

Laroche F. (2007). *Contribution à la sauvegarde des objets techniques anciens par l'archéologie industrielle avancée. Proposition d'un Modèle d'information de référence muséo-*

logique et d'une Méthode interdisciplinaire pour la Capitalisation des connaissances du Patrimoine. Thèse de doctorat soutenue à l'École Centrale de Nantes.

Leduc J. (1999). « Découper le temps ». In Leduc J. (dir.), *Les historiens et le temps. Conceptions, problématiques, écritures*. Paris : Seuil, pp. 91-133.

Malraux A. (1965). *Le musée imaginaire*. Paris : Folio Essais.

Mason M. (2015). « Prototyping practices supporting interdisciplinary collaboration in digital media design for museums ». *International Journal of Museum Management and Curatorship*, 30(5), pp. 394-426.

de Massary X. & Coste G. (2007). « Principes, méthodes et conduite de l'inventaire général du patrimoine culturel ». *Documents & Methodes*. Paris : Ministère de la Culture et de la Communication, p. 9.

Mendes P. N., Jakob M. & Bizer C. (2012). « DBpedia for NLP : A Multilingual Cross-domain Knowledge Base ». In Piperidis N. C. (dir.). *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey : European Language Resources Association (ELRA).

Park H. (2014). *Crowdsourcing Structured Data*. Thèse de doctorat soutenue à Stanford InfoLab.

Patel M. *et al.* (2005). « Metadata requirements for digital museum environments ». *International Journal on Digital Libraries*, 5(3), pp. 179-192.

Prévôt N. (2013). « The Digital Puzzle of the talatat from Karnak. A Tool for the Three-Dimensional Reconstruction of Theban Buildings from the Reign of Amenhotep IV ». *Texts, Languages & Information Technology in Egyptology*, pp. 129-138.

Quantin M. (2015). « Représentation spatiale des connaissances : Connaissances et représentations spatiales, quelles dépendances ? ». *Formation DG Patrimoines*, 173.

Rosenzweig R. (2016). *Omeka*. CMS Web. Center for History and New Media, George Mason University, Fairfax [En ligne].

Shneiderman B. (1999). « The Eyes Have It : a Task by Data Type Taxonomy for Information Visualizations ». *Proceedings 1996 IEEE Symposium on Visual Languages*, pp. 336-343.

Srinivasan R. & Huang J. (2005). « Fluid ontologies for digital museums ». *International Journal on Digital Libraries*, 5(3), pp. 193-204.

Stiegler B. (2009). « Technologies culturelles et économie de la contribution ». *Culture et recherche*, (121), Paris : Ministère de la Culture et de la Communication, p. 31.

TICCIH (2003). *The Nizhny Tagil Charter for the Industrial Heritage*, pp. 1-6.

UNESCO (2003). *Charte sur la conservation du patrimoine numérique*, pp. 1-5.

Vadelorge L. (2014). « Inventaire Général et recherche, quelles interactions ? ». *L'Observatoire* (45).

Veyne P. (1971). *Comment on écrit l'Histoire*. Paris : Seuil.

Welger-Barboza C. (2001). *Du musée virtuel au musée médiathèque: le patrimoine à l'ère du document numérique*. Paris : l'Harmattan.

Zacklad M. (2007). « Une théorisation communicationnelle et documentaire des TIC ». *Humanités numériques 2 : socio-informatique et démocratie cognitive*. Hermes Science Publications, pp. 20-35.